Toutenburg, Srivastava:

A Revisit to the Application of Weighted Mixed Regression Estimation in Linear Regression Models with Missing Data

Sonderforschungsbereich 386, Paper 241 (2001)

Online unter: http://epub.ub.uni-muenchen.de/

A Revisit to the Application of Weighted Mixed Regression Estimation in Linear Regression Models with Missing Data

H. Toutenburg
Institut für Statistik, Universität München
80799 München, Germany

V. K. Srivastava
Department of Statistics, lucknow University
Lucknow-226 007, India

May 23, 2001

Abstract

This paper deals with the application of the weighted mixed regression estimation of the coefficients in a linear model when some values of some of the regressors are missing. Taking the weight factor as an arbitrary scalar, the performance of weighted mixed regression estimator in relation to the conventional least squares and mixed regression estimators is analyzed and the choice of scalar is discussed. Then taking the weight factor as a specific matrix, a family of estimators is proposed and its performance properties under the criteria of bias vector and mean squared error matrix are analyzed.

1 Introduction

In many practical applications, the available data for regression analysis is found to contain some missing values. Consequently, the standard statistical procedures for deducing the inferences pertaining to the parameters of the model cannot be applied: see, e. g. Little and Rubin (1987) for an interesting account. Now there are two alternatives. One is to follow the amputation procedure which involves discarding the incomplete observations and using the complete observations only. Contending that the deleted observations may contain some valuable information, the other alternative is to adopt some imputation procedure which consists of finding substitutes for the missing values and employing the thus obtained data set. Both the alternatives have their own limitations and qualifications, and it is generally hard to achieve uniform superiority of amputation over imputation or vice-versa.

When some values of some of the regressors in a linear regression model are missing, there are several procedures for finding the imputed values of missing observations; see, e. g. Little (1992), Little and Rubin (1987) and Rao and Toutenburg (1999). Employing the imputed values to fill up the missing values, Toutenburg, Heumann, Fieger and Park (1995) have considered the estimation of regression coefficients under a mixed regression framework. The conventional mixed regression estimator for the coefficient vector is presented and its superiority over the least squares estimator which ignores the incomplete observations is discussed under several criteria based on the mean squared error matrix. They have also considered the application of weighted mixed regression estimation in which possibly unequal weights are given to the complete and repaired parts of the data set with the help of a scalar. Instead of a scalar the choice of which is arbitrary and is a matter of practitioner's preference, we propose the assignment of weights with the help of specific matrices. Based on such a proposition, a class of estimators is presented and its efficiency properties are analyzed.

The plan of this article is as follows. In Section 2, we describe the model and present the conventional estimators for the vector of regression coefficients along with their efficiency properties. In Section 3, we present the weighted mixed regression estimator considered by Toutenburg et al. (1995) and discuss its performance properties. An alternative estimator arising from the application of weighted mixed regression estimation using a different weighting procedure is then considered in Section 4 and a family of estimators termed as t-class is proposed. The properties are also discussed using the large sample asymptotic theory. Some concluding remarks are finally offered in Section 5. Lastly, the Appendix gives the derivation of results stated in the Theorem.

2 Model Specification and the Conventional Estimators

Let us postulate the following linear regression model with n complete observations:

$$y_c = X_c\beta + \epsilon_c \tag{2.1}$$

where y_c denotes a $n \times 1$ vector of n observations on the study variable, X_c is a $n \times K$ full column rank matrix of n observations on the K regressors of the model, β is a $K \times 1$ vector of the coefficients associated with them and ϵ_c is a $n \times 1$ vector of disturbances.

Further, we assume the availability of m incomplete observations in which some values of some of the regressors are missing. Thus, if y_* denotes the $m \times 1$ vector of m observations on the study variable, X_* is the $m \times K$ matrix with not necessarily of full column rank of m observations on the K regressors such that every row of it has at least one missing value and ϵ_* denotes the $m \times 1$ vector of disturbances, we have

$$y_* = X_*\beta + \epsilon_* \, . \tag{2.2}$$

It is assumed that the elements of ϵ_c and ϵ_* are independently and identically distributed with mean 0 and finite but unknown variance σ^2.

When the incomplete observations are totally dropped from the data set, the application of least squares procedures to (2.1) gives the following estimator of β:

$$b_c = (X - c' X_c)^{-1} X_c' y_c \tag{2.3}$$

which is unbiased with variance covariance matrix as

$$\begin{aligned} V(b_c) &= E(b_c - \beta)(b_c - \beta)' \\ &= \sigma^2 (X_c' X_c)^{-1} \\ &= \sigma^2 S_c \quad (\text{say})\,. \end{aligned} \tag{2.4}$$

Alternatively, one may employ some imputation procedure for completing the data set; see, e. g., Little (1992), Little and Rubin (1987) and Rao and Toutenburg (1999) for an interesting account. Following some imputation procedures and inserting the imputed values in place of the missing values in X_*, suppose that we obtain the matrix X_R. Thus we can write (2.2) as follows:

$$y_* = X_R \beta + v_* \,;\, v_* = (\epsilon_* + \sigma Z \beta) \tag{2.5}$$

where

$$Z = \frac{1}{\sigma}(X_* - X_R)\,. \tag{2.6}$$

Combining (2.1) and (2.5) to form the mixed regression framework, the application of least squares provides the following estimator:

$$b_R = (X_c' X_c + X_R' X_R)^{-1} (X_c' y_c + X_R' Y_*) \tag{2.7}$$

which is generally biased with bias vector

$$\begin{aligned} B(b_R) &= E(b_R - \beta) \\ &= \sigma (X_c' X_c + X_R' X_R)^{-1} X_R' Z \beta \end{aligned} \tag{2.8}$$

and mean squared error matrix as

$$\begin{aligned} MSE(b_R) &= E(b_R - \beta)(b_R - \beta)' \\ &= \sigma^2 [S + S X_R' Z \beta \beta' Z' X_R S] \end{aligned} \tag{2.9}$$

where

$$S = (X_C' X_c + X_R' X_R)^{-1}\,. \tag{2.10}$$

Toutenburg et al. (1995) have compared the estimators b_c and b_R under the criteria of mean squared error matrix, total mean squared error (trace of mean squared error matrix) and predictive mean squared error (trace of $X_c' X_c$ times the mean squared error matrix), and have deduced conditions for the superiority of b_R over b_c. For instance, a necessary and sufficient condition for the variance covariance matrix of b_c to exceed the mean squared error matrix of b_R by a non-negative definite matrix is

$$\beta' Z' (I + X_R S_c X_R')^{-1} Z \beta \leq 1\,. \tag{2.11}$$

A procedure for testing this condition is also outlined by them.

3 The Weighted Mixed Regression Estimator

Toutenburg et al. (1995) have considered the minimization of the quantity

$$(y_c - X_c\beta)'(y_c - X_c\beta) + \lambda(y_+ - X_R\beta)'(y_+ - X_R\beta) \tag{3.1}$$

with respect to the elements of β and have found the following weighted mixed regression estimator of β:

$$b_\lambda = (X_c'X_c + \lambda X_R'X_R)^{-1}(X_c'y_c + \lambda X_R'y_*) \tag{3.2}$$

where λ is a positive scalar not exceeding 1 and can be regarded as reflecting the weight being assigned to the repaired part of the data set in relation to the complete observations.

Viewing (3.2) as a family of estimators characterized by the scalar λ, it is interesting to note that b_λ reduces to b_c for $\lambda = 0$ and to b_R for $\lambda = 1$.

Assuming λ to be nonstochastic, it is easy to see that the bias vector of b_λ is given by

$$\begin{aligned} B(b_\lambda) &= E(b_{\lambda-\beta}) \\ &= \sigma\lambda(X_c'X_c + \lambda X_R'X_R)^{-1}X_R'Z\beta \end{aligned} \tag{3.3}$$

while its mean squared error matrix is

$$\begin{aligned} MSE(b_\lambda) &= E(b_{\lambda-\beta})(b_{\lambda-\beta})' \\ &= \sigma^2 S_\lambda - \sigma^2\lambda S_\lambda X_R'[(1-\lambda)I + \lambda Z\beta\beta'Z']X_R S_\lambda \end{aligned} \tag{3.4}$$

where

$$S_\lambda = (X_c'X_c + \lambda X_R'X_R)^{-1}\,. \tag{3.5}$$

These expressions are presented by Toutenburg et al. (1995) but no comparison with the corresponding expressions for b_c and b_R is reported. Let us fill up this gap.

Comparing b_c, b_R and b_λ, we observe that b_c is an unbiased estimator of β while b_R and b_λ are generally biased.

For the comparison of bias, we take the criterion as the length of the bias vector wich may furnish an idea about magnitude of bias. Under this criterion, it is seen from (2.8) and (3.3) that b_λ has smaller length of bias vector in comparison to b_R when

$$\frac{\lambda^2\beta'Z'X_R S_\lambda^2 X_R'Z\beta}{\beta'Z'X_R S^2 X_R Z\beta} \le 1\,. \tag{3.6}$$

As the matrix $[(X_c'X_c + \lambda X_R'X_R) - \lambda(X_c'X_c + X_R'X_R)]$ is positive definite for $0 < \lambda < 1$, the matrix

$$[(X_c'X_c + X_R'X_R)^{-1} - \lambda(X_c'X_c + \lambda X_R'X_R)^{-1}] = S - \lambda S_\lambda$$

is also positive definite. Using it, we see that the condition (3.6) always holds true. Thus the weighted mixed regression estimator has smaller length of bias vector in comparison to the conventional mixed regression estimator.

If we write

$$G = (I + \lambda X_R S_c X_R')^{-1} \tag{3.7}$$

and use the identity

$$S_\lambda = (X_c' X_c + \lambda X_R' X_R)^{-1} = S_c - \lambda S_c X_R' G X_R S_c \tag{3.8}$$

in (3.4), we get

$$MSE(b_\lambda) = \sigma^2 S_c - \sigma^2 \lambda S_c X_R' G[G^{-1} + (1-\lambda)I - \lambda Z\beta\beta' Z']G X_R S_c \,. \tag{3.9}$$

Comparing it with (2.4), it is observed that th estimator b_λ is superior to b_c according to the criterion of mean squared error matrix when the matrix expression

$$\begin{aligned} & G^{-1} + (1-\lambda)I - \lambda Z\beta\beta' Z' \\ = \; & (2-\lambda)I + \lambda X_R S_c X_R' - \lambda Z\beta\beta' Z' \end{aligned}$$

is non-negative definite for which a necessary and sufficient condition is

$$\beta' Z' [(\frac{2-\lambda}{\lambda})I + X_R S_c X_R']^{-1} Z\beta \leq 1 \,. \tag{3.10}$$

As $0 \leq \lambda < 1$, a comparison of (3.10) with the condition (2.11) reveals that the weighted mixed regression estimator b_λ dominates the unbiased estimator b_c over a relatively wider range of situations when compared with the dominance of the conventional mixed regression estimator b_R over b_c.

Finally, let us compare the two biased estimators b_R and b_λ.

Writing

$$D = [I + (1-\lambda) X_R S_\lambda X_R']^{-1} \tag{3.11}$$

we can express

$$\begin{aligned} (X_c' X_c + X_R' X_R)^{-1} &= [S_\lambda^{-1} + (1-\lambda) X_R' X_R]^{-1} \\ &= S_\lambda - (1-\lambda) S_\lambda X_R' D X_R S_\lambda \,. \end{aligned} \tag{3.12}$$

Substituting it in (2.9), we get

$$MSE(b_R) = \sigma^2 S_\lambda - \sigma^2 S_\lambda X_R' D[(1-\lambda)D^{-1} - Z\beta\beta' Z'] D X_R S_\lambda \,. \tag{3.13}$$

Comparing it with (3.4), we observe that

$$MSE(b_\lambda) - MSE(b_R) = \sigma^2 S_\lambda X_R' (A - DZ\beta\beta' Z' D) X_R S_\lambda \tag{3.14}$$

where

$$A = (1-\lambda)(D-\lambda I) + \lambda^2 Z\beta\beta' Z' . \tag{3.15}$$

Thus the mean squared error matrix of b_λ exceeds the mean squared error matrix of b_R by a non-negative definite matrix when $(A - DZ\beta\beta' Z' D)$ is a non-negative definite matrix for which a necessary and sufficient condition is

$$\beta' Z' D A^{-1} DZ\beta \leq 1 . \tag{3.16}$$

This condition is satisfied so long as

$$\beta' Z' D(D-\lambda I)^{-1} DZ\beta \leq (1-\lambda)$$

or

$$\beta' Z'[I + (1-2\lambda) X_R S_\lambda X_R' - \lambda(1-\lambda) X_R S_\lambda X_R' X_R S_\lambda X_R']^{-1} Z\beta \leq 1 . \tag{3.17}$$

On the other hand, for the dominance of b_λ over b_R, consider the difference

$$MSE(b_R) - MSE(b_\lambda) = \sigma^2 S_\lambda X_R'(DZ\beta\beta' ZD - A) X_R S_\lambda . \tag{3.18}$$

As the matrix A is positive definite, the matrix expression inside the brackets on the right hand side of (3.18) can never be non-negative definite except in the trivial case $m = 1$. In other words, the weighted mixed regression estimator b_λ does not dominate the conventional mixed regression estimator b_R with respect to the criterion of mean squared error matrix so long as the number of incomplete observations is more than one.

Lastly, let us consider he optimal choice of the scalar λ. It is obvious from (3.3) that non non-zero value of λ can be found for which the estimator b_λ is unbiased or approximately unbiased. Similarly, it is seen from (3.4) that we cannot choose a value of λ for which the expression of the mean squared error matrix attains its minimum. Even when we attempt to minimize the total mean squared error, i.e., the trace of the mean squared error matrix, the optimal choice of λ is provided by the solution of equation

$$\begin{aligned} &\mathrm{tr}(X_c' X_c + \lambda^2 X_R' X_R)(X_c' X_c + \lambda X_R' X_R)^{-2} \\ &\quad + \lambda^2 \beta' Z' X_R (X_c' X_c + \lambda X_R' X_R)^{-2} X_R' Z\beta = 0 . \end{aligned} \tag{3.19}$$

As this is a typically nonlinear equation, it is difficult to deduce a neat expression for the solution except in the particular case of $m = 1$; see Toutenburg et al. (1995, Section 4.2). Pursuing the case of $m = 1$, they have proposed to replace $(X_+ - X_R)\beta$ and σ^2 in the optimum value of λ by their unbiased estimators and have reported some simulation results.

4 An Alternative Weighted Mixed Regression Estimator

Considering the model equations (2.1) und (2.5), we observe that

$$[E(\epsilon_c \epsilon_c')]^{-1} = \frac{1}{\sigma^2} I \tag{4.1}$$

$$
\begin{aligned}
[E(v_* v_*')]^{-1} &= \frac{1}{\sigma^2}(I + Z\beta\beta' Z')^{-1} \\
&= \frac{1}{\sigma^2}(I - \frac{1}{1-\beta' Z' Z\beta} Z\beta\beta' Z') \\
&= \frac{1}{\sigma^2}\Lambda \quad \text{(say)}\,.
\end{aligned} \tag{4.2}
$$

Instead of the weighted sum (3.1), let us take the criterion function as the following weighted sum

$$
(y_c - X_c\beta)'[E(\epsilon_c\epsilon_c')]^{-1}(y_c - X_c\beta) + (y_* - X_R\beta)'[E(v_* v_*')]^{-1}(y_* - X_R\beta)
$$

or equivalently

$$
(y_c - X_c\beta)'(y_c - X_c\beta) + (y_* - X_R\beta)'\Lambda(y_* - X_R\beta)\,. \tag{4.3}
$$

Minimizing it with respect to the elements of β, we obtain the following estimator of β:

$$
\tilde{\beta} = (X_c' X_c + X_R'\Lambda X_R)^{-1}(X_c' y_c + X_R'\Lambda y_*)\,. \tag{4.4}
$$

Using the identity

$$
\begin{aligned}
&(X_c' X_c + X_R'\Lambda X_R)^{-1} \\
&= (X_c' X_c + X_R' X_R - \frac{1}{1+\beta' Z' Z\beta} X_R' Z\beta\beta' Z' X_R)^{-1} \\
&= (X_c' X_c + X_R' X_R)^{-1} \\
&\quad +[1 + \beta' Z' Z\beta - \beta' Z' X_R (X_c' X_c + X_R' X_R)^{-1} X_R' Z\beta]^{-1} \\
&\quad \cdot (X_c' X_c + X_R' X_R)^{-1} X_R' Z\beta\beta' Z' X_R (X_c' X_c + X_R' X_R)^{-1}
\end{aligned}
$$

we can express

$$
\begin{aligned}
\tilde{\beta} = b_R - \\
\frac{\beta' Z'(y_+ - X_R b_R)}{1 + \beta' Z'[I - X_R(X_c' X_c + X_R' X_R)^{-1} X_R'] Z\beta}(X_c' X X_c + X_R' X_R)^{-1} X_R' Z\beta\,.
\end{aligned} \tag{4.5}
$$

As this estimator involves σ^2 and $(X_+ - X_R)\beta$ which are unknown, we propose to replace them by their unbiased estimators s_c^2 and $y_+ - X_R b_c$ with

$$
s_c^2 = (\frac{1}{n-K})(y_c - X_c b_c)'(y_c - X_c b_c)\,. \tag{4.6}
$$

Such a proposition leads to the following feasible version of $\tilde{\beta}$:

$$
\tilde{\beta}_F = b_R - h(X_c' X_c + X_R' X_R)^{-1} X_R'(y_* - X_R b_c) \tag{4.7}
$$

where

$$
h = \frac{(y_* - X_R b_c)'(y_* - X_R b_R)}{s_c^2 + (y_* - X_R b_c)'[I - X_R(X_c' X_c + X_R' X_R)^{-1} X_R'](y_* - X_R b_c)}\,. \tag{4.8}
$$

Looking at the form of the feasible estimator (4.7), we propose the following t-class of estimators for β:

$$\hat{\beta}_t = b_R - th(X_c'X_c + X_R'X_R)^{-1}X_R'(y_* - X_R b_c) \tag{4.9}$$

where t is any non-negative scalar characterizing the etimator.

Observing that

$$X_R'(y_* - X_R b_c) = (X_c'y_c + X_R'y_*) - (X_c'X_c + X_R'X_R)b_c$$

we can express

$$\hat{\beta}_t = thb_c + (1 - th)b_R\,. \tag{4.10}$$

Thus $\hat{\beta}_t$ is a convex linear combination of the conventional least squares and mixed regression estimators.

The exact expressions for the bias vector and the mean squared error matrix of $\hat{\beta}_t$ cannot be derived unless we specify the distribution of disturbances. Even when we make a simplifying assumption like normal distribution, it can be well appreciated that the exact expressions will have sufficiently intricate form and will not permit us to draw any meaningful and clear inference. We therefore consider their large sample approximations without assuming any specific distribution for the disturbances in the model.

For the application of the large sample asymptotic theory, we assume that the regressors are asymptotically cooperative so that the limiting forms of the matrices $n^{-1}X_c'X_c$,$m^{-1}X_*'X_*$,$m^{-1}X_R'X_R$ and $m^{-1}X_R'X_*$ as m and n grow large are finite. Further, it is assumed that the limiting form of $n^{-1}X_c'X_c$ is also nonsingular.

Let us write

$$\begin{aligned} \theta &= \frac{1}{m}\beta'Z'(I - X_R S X_R')Z\beta \\ \alpha &= (1+\theta)^{-1}(\frac{m-1}{m} + \theta) \end{aligned} \tag{4.11}$$

The following results are derived in Appendix.

<u>THEOREM:</u> Assuming the characterizing scalar t to be nonstochastic, the large sample asymptotic approximations for the bias vector and the mean squared error matrix of the t-class estimators of β to order $O(n^{-1})$ are given by

$$\begin{aligned} B(\hat{\beta}_t) &= E(\hat{\beta}_t - \beta) \\ &= \sigma(1 - \alpha t)SX_R'Z\beta \end{aligned} \tag{4.12}$$

$$\begin{aligned} MSE(\hat{\beta}_t) &= E(\hat{\beta}_t - \beta)(\hat{\beta}_t - \beta)' \\ &= \sigma^2[t^2 S_c + (1 - t^2)S] \\ &\quad + \sigma^2(1-t)[1 + (1-2\alpha)t]SX_R'Z\beta\beta'Z'X_R S\,. \end{aligned} \tag{4.13}$$

4.1 Comparison of Bias Vectors

It is clearly seen from (2.8), (3.3) and (4.12) that all the three estimators b_R, b_λ and $\hat{\beta}_t$ are generally biased. However, $\hat{\beta}_t$ becomes unbiased to order $O(n^{-1})$, e. g., approximately unbiased if we choose $t = (1/\alpha)$ which is always larger than one as α lies between 0 and 1 by virtue of its specification.

For studying the magnitude of bias, let us take the length of bias vector as the measure. With respect to such a measure, we observe from (2.8) and (4.12) that $\hat{\beta}_t$ is superior to b_R when

$$0 < t < \frac{2}{\alpha} . \tag{4.14}$$

TZhe reverse is true, i. e., $\hat{\beta}_t$ fails to beat b_R when t exceeds $(2/\alpha)$.

Similarly, it is seen from (3.3) and (4.12) that $\hat{\beta}_t$ is better than b_λ when

$$\frac{\beta' Z' X_R S_\lambda^2 X_R' Z\beta}{\beta' Z' X_R S X_R' Z\beta} > (\frac{1-\alpha t}{\lambda})^2 . \tag{4.15}$$

As $(S_\lambda - S)$ is a non-negative definite matrix, the condition (4.15) is satisfied so long as

$$(\frac{1-\alpha t}{\lambda})^2 < 1$$

or

$$(\frac{1-\lambda}{\alpha}) < t < (\frac{1+\lambda}{\alpha}) . \tag{4.16}$$

The opposite is true, i. e., $\hat{\beta}_t$ is inferior to b_λ as long as

$$t > \frac{2}{\alpha} \tag{4.17}$$

where use has been made of the non-negative definiteness of the matrix $(S - \lambda S_\lambda)$.

As λ is assumed to lie between 0 and 1, it is observed from (4.14) and (4.16) that the t-class estimators have smaller length of bias vector, to the given order of approximation, in comparison to the estimators b_R and b_λ provided that the characterizing scalar t is chosen according to the constraint (4.16).

4.2 Comparison of Mean Squared Matrices

Next, let us compare the estimators with respect to the criterion of mean squared error matrix.

First comparing $\hat{\beta}_t$ with the unbiased estimator b_c, we observe from (2.4) and (4.13) that

$$\begin{aligned} V(b_c) - MSE(\hat{\beta}_t) &= \sigma^2(1-t^2)\cdot \\ &\quad [(S_c - S) - (1 - \frac{2\alpha t}{1+t})SX'_R Z\beta\beta' Z' X_R S] \\ &= SX'_R \Delta_c X_R S \end{aligned} \quad (4.18)$$

where

$$\Delta_c = (1-t^2)[(I + X_R S_c X'_R) - (1 - \frac{2\alpha t}{1+t})Z\beta\beta' Z']\,. \quad (4.19)$$

Clearly, the matrix Δ_c is non-negative definite implying the superiority of $\hat{\beta}_t$ over b_c under mean squared error matrix if and only if

$$(1 - \frac{2\alpha t}{1+t})\beta' Z'(I + X_R S_c X'_R)Z\beta \leq 1 \quad (4.20)$$

provided that the characterizing scalar t does not exceed one.

It is interesting to note that the condition (4.20) will always be true as long as

$$\alpha > (\frac{1+t}{2t})\,; \quad 0 < t < 1\,. \quad (4.21)$$

On the other hand, the unbiased estimator b_c remains superior to $\hat{\beta}_t$ when t is greater than one and either of the following two conditions is satisfied:

$$\alpha > (\frac{1+t}{2t}) \quad (4.22)$$

$$(1 - \frac{2\alpha t}{1+t})\beta' Z'(I + X_R S_c X'_R)Z\beta \leq 1\,. \quad (4.23)$$

Next, from (2.9) and (4.13), we have

$$MSE(b_R) - MSE(\hat{\beta}_t) = -\sigma^2 t^2 SX'_R \Delta_R X_R S \quad (4.24)$$

where

$$\Delta_R = (I + X_R S_c X'_R) + [1 + 2\alpha(\frac{1-t}{t})]Z\beta\beta' Z'\,. \quad (4.25)$$

We thus see, that $\hat{\beta}_t$does not dominate b_R for $m > 1$. On the other hand, b_R dominates $\hat{\beta}_t$ as long as t does not exceed one. For t exceeding one, the necessary and sufficient condition is

$$[2\alpha(\frac{t-1}{t}) - 1]\beta' Z'(I + X_R S_c X'_R)^{-1} Z\beta \leq 1\,. \quad (4.26)$$

When we compare b_λ and $\hat{\beta}_t$ under the criterion of mean squared error matrix, it may be appreciated from expression (3.4) and (4.13) that it is difficult to deduce necessary and sufficient conditions for the superiority of b_λ over $\hat{\beta}_t$ and vice-versa.

4.3 Comparison of Weighted Mean Squared Errors

Let us take the performance criterion as the weighted mean squared error defined by the trace of the mean squared error matrix multiplied by the matrix $S^{-1} = (X_c'X_c + X_R'X_R)$.

From (2.4), (2.9), (3.4) and (4.13), we observe that the weighted mean squared errors of the estimators b_c, b_R, b_λ and $\hat{\beta}_t$ can be expressed as

$$\begin{aligned}
WMSE(b_c) &= E(b_c - \beta)'S^{-1}(b_c - \beta) && (4.27)\\
&= \sigma^2 \operatorname{tr} S^{-1}S_c\\
WMSE(b_R) &= E(b_R - \beta)'S^{-1}(b_R - \beta) && (4.28)\\
&= \sigma^2[K + \beta'Z'X_RSX_R'Z\beta]\\
WMSE(b_\lambda) &= E(b_\lambda - \beta)'S^{-1}(b_\lambda - \beta) && (4.29)\\
&= \sigma^2[\lambda \operatorname{tr} S_\lambda S^{-1} + (1-\lambda)\operatorname{tr} S_cS_\lambda S^{-1}S_\lambda\\
&\quad -\lambda^2\beta'Z'X_RS_\lambda S^{-1}S_\lambda X_R'Z\beta]\\
WMSE(\hat{\beta}_t) &= E(\hat{\beta}_t - \beta)'S^{-1}(\hat{\beta}_t - \beta) && (4.30)\\
&= \sigma^2[t^2 \operatorname{tr} S^{-1}S_c + (1-t^2)K\\
&\quad +(1-t^2)(1-\frac{2\alpha t}{1+t^2})\beta'Z'X_RSX_R'Z\beta]
\end{aligned}$$

Writing

$$\delta = \frac{\beta'Z'X_RSX_R'Z\beta}{(\operatorname{tr} S^{-1}S_c - K)} \tag{4.31}$$

it is seen from (4.27) and (4.30) that the t-class estimators are superior to the estimator b_c under the criterion of weighted mean squared error when either of the following conditions is satisfied:

$$(1-\frac{2\alpha t}{1+t})\delta > 1\,; \quad t > 1 \tag{4.32}$$

$$(1-\frac{2\alpha t}{1+t})\delta < 1\,; \quad t < 1\,. \tag{4.33}$$

Next, we observe from (4.28) and (4.30) that the t-class estimators are better than b_R when we choose t according to the following constraint

$$[1 + 2\alpha(\frac{1-t}{t})]\delta > 1\,. \tag{4.34}$$

Similarly, it follows from (4.29) and (4.30) that $\hat{\beta}_t$ is better than b_λ when the expression

$$\begin{aligned}
&\lambda \operatorname{tr} S_\lambda S^{-1} + (1-\lambda)\operatorname{tr} S_cS_\lambda S^{-1}S_\lambda - t^2 \operatorname{tr} S^{-1}S_c - (1-t^2)K && (4.35)\\
&\quad -\beta'Z'X_R[\lambda^2 S_\lambda S^{-1}S_\lambda + (1-t^2)(1-\frac{2\alpha t}{1+t})S]X_R'Z\beta
\end{aligned}$$

is positive.

As the matrix $(S - \lambda S_\lambda)$ is non-negative definite, the expression (4.35) is positive so long as

$$[2(1-\alpha t) - (1-2\alpha)t^2]\delta < [\frac{\lambda \operatorname{tr} S_\lambda S^{-1} + (1-\lambda) S_c S_\lambda S^{-1} S_\lambda - K}{\operatorname{tr} S^{-1} S_c - K} - t^2] \tag{4.36}$$

provided that the quantity on the right hand side of the inequality is positive.

If we minimize the expression (4.30) with respect to t, the optimum value of t is given by

$$t_o = \frac{\alpha\delta}{1-(1-2\alpha)\delta} \tag{4.37}$$

provided that $(1-2\alpha)\delta$ is less than one.

As α and δ involve unknown quantities σ^2 and $(X_* - X_R)\beta$, we may replace them by their unbiased estimators. This proposition leads to the following estimators of α and δ:

$$\hat{\alpha} = 1 - [m + \frac{1}{s_c^2}(y_* - X_R b_c)'(I - X_R S X_R')(y_* - X_R b_c)]^{-1} \tag{4.38}$$

$$\hat{\delta} = \frac{(Y_* - X_R b_c)' X_R S X_R' (Y_* - X_R b_c)}{s_c^2(\operatorname{tr} S^{-1} S_c - K)} \tag{4.39}$$

which when substituted in (4.37) provides a feasible version of t_o, viz.;

$$\hat{t}_o = \frac{\hat{\alpha}\hat{\delta}}{1-(1-2\hat{\alpha})\hat{\delta}} \tag{4.40}$$

that can be readily used in practice.

Other feasible versions of t_o can be constructed by using b_R or $\hat{\beta}_1$ (i. e., $\hat{\beta}_t$ with $t = 1$) in place of b_c in (4.38) and (4.40).

5 Some Concluding Remarks

Employing the mixed regression framework, we have considered the estimation of the vector of coefficients in a linear regression model when some values of some of the regressors are missing. Following (nicht in LitDB!)the minimization of a weighted combination of the two sums of squares of errors arising from complete observations and repaired observations has been considered by Toutenburg et al. (1995) taking the weight factor as an arbitrary scalar. The resulting weighted mixed regression estimator essentially defines a class of estimators characterized by the scalar, and the conventional least squares and mixed regression estimators are found to be particular cases of it. The weighted mixed regression estimator is seen to possess smaller length of bias vector in comparison to the conventional mixed regression estimator. Further, taking the performance criterion to be the mean squared error matrix and comparing the necessary and sufficient conditions for superiority of the conventional mixed regression estimator and the weighted mixed regression estimator over the unbiased least squares estimator, it is observed that the weighted mixed regression estimator performs better than the least squares estimator over a relatively wider range of situations when compared with the situations under which the conventional mixed regression estimator is better than the least squares estimator. However, when the two biased estimators are compared, the weighted mixed regression estimator fails to dominate the conventional mixed regression estimator under mean squared error matrix criterion. For the converse of it, i. e., the dominance of the conventional mixed regression estimator over the weighted mixed regression estimator, we succeed in deducing a necessary and sufficient condition.

So far as the optimal choice of characterizing scalar in the weighted mixed regression estimator is concerned, it is hard to determine. For instance, no value of the scalar can be specified for which the expression for the mean squared error matrix attains its minimum. even when the criterion is taken to be a scalar based on the mean squared error matrix, the determination of the optimal values requires solving a typically non-linear equation and no neat algebraic expression can be found. Consequently, iterative procedures are required to find it m´numerically but then the properties of the resulting estimator cannot be investigated analytically. Besides it, the non-linear equation contains some unknown quantities. However, one can replace them by theri estimators as recommended by Toutenburg et al. (1995).

Instead of using a scalar weight for combining the two sums of squares of errors, we propose to use a specific matrix based on the second order moment matrices of the disturbances in the models associated with the complete and repaired observations. This proposition leads to an estimator which involves some unknown quantities. Replacing them by their unbiased estimators in a bid to obtain a feasible version and introducing a characterizing scalar t, we have presented a family of estimators which we term as t-class estimators. These estimators turn out to be the convex linear combinations of the conventional least squares and mixed regression estimators.

Analyzing the performance properties of the t-class estimaotrs with the help of large sample asymptotic theory, it is found that the t-class estimators are generally biased.

However, the bias in large samples may vanish through a suitable choice of the characterizing scalar t. Examining the length of the bias vector to the given order of approximation, we have determined the range of t in which the t-class estimators have smaller magnitude of the length of bias vector in comparison to the conventional mixed regression and weighted mixed regression estimators.

Comparing the estimators with respect to the criterion of mean squared error matrix, it is observed that the t-class estimators with $0 < t < 1$ dominate the unbiased least squares estimator under certain condition. However, the t-class estimators fail to dominate the mixed regression estimator when the number of incomplete observations is more than one. On the contrary, the mixed regression estimator dominates the t-class estimator with t less than one. This result remains true for t greater than one when a certain condition is satisfied.

Taking the performance criterion as the weighted mean squared error, we have worked out the conditions under which the t-class estimators are superior to the least squares, mixed regression and weighted mixed regression estimators. A value of the characterizing scalar t is also determined which minimizes the weighted mean squared error. Such a value is found to contain some unknown quantities, and accordingly feasible versions are presented.

Our investigations are fairly general though asymptotic in nature as no assumption regarding the distribution of disturbances like normality is made. However, it will be interesting to analyze the performance properties in finite samples assuming some specific distribution for the disturbances. Perhaps an elaborate numerical study based on simulation and/or bootstrap methodologies may provide some useful findings.

Appendix

Besides the quantities defines in (4.), let us write

$$\begin{aligned}
Q_{cc} &= \frac{1}{n}X_c'X_c\,, \quad Q_{**} = \frac{1}{m}X_*'X_*\,, \\
Q_{RR} &= \frac{1}{m}X_R'X_R\,, \quad Q_{R*} = \frac{1}{m}X_R'X_*\,, \\
Q &= (Q_{cc} + fQ_{RR})^{-1}\,, \quad \phi = \frac{1}{\sigma}(Q_{R*} - Q_{RR})\beta\,, \\
\theta &= \frac{1}{\sigma^2}\beta'(Q_{**} + Q_{RR} - Q_{R*} - Q_{R*}')\beta - f^2\phi'Q\phi\,, \\
u_c &= \frac{1}{n^{1/2}}X_c'\epsilon_c\,, \quad w_c = n^{1/2}(\frac{\epsilon_c'\epsilon_c}{n} - \sigma^2)\,, \\
u_* &= \frac{1}{m^{1/2}}X_*'\epsilon_*\,, \quad w_* = m^{1/2}(\frac{\epsilon_*'\epsilon_*}{m} - \sigma^2)\,, \\
u_R &= \frac{1}{m^{1/2}}X_R'\epsilon_*\,, \quad f = \frac{m}{n}
\end{aligned}$$

so that $Q_{cc}, Q_{**}, Q_{RR}, Q_{R*}, Q, \phi, \theta$ and f are of order $O(1)$ while u_c, u_*, u_R, w_c and w_* are of order $O_p(1)$.

Thus we can express

$$(b_c - \beta) = \frac{1}{n^{1/2}}Q_{cc}^{-1}u_c \tag{A.1}$$

$$(b_R - \beta) = fQ\delta + \frac{1}{n^{1/2}}Q(u_c + f^{1/2}u_r)\,. \tag{A.2}$$

Next, we observe that

$$\begin{aligned}
(y_c - X_c b_c) &= \epsilon_c - \frac{1}{n^{1/2}} X_c Q_{cc}^{-1} u_c \\
(y_* - X_R b_c) &= \epsilon_* + (X_* - X_R)\beta - \frac{1}{n^{1/2}} X_R Q_{cc}^{-1} u_c \\
(y_* - X_R b_R) &= \epsilon_* + (X_* - X_R)\beta - \sigma f X_R Q \phi - \frac{1}{n^{1/2}} X_R Q (u_c + f^{1/2} u_R) \\
\frac{1}{n} Q X_R' X_R Q_{cc}^{-1} &= (Q_{cc}^{-1} - Q) \\
s_c^2 &= \sigma^2 + \frac{w_c}{n^{1/2}} + O_p(n^{-1})\,.
\end{aligned}$$

Using these, it is easy to see that

$$\frac{1}{n}(y_* - X_R b_c)'(y_* - X_R b_R) = \sigma^2 f(1+\theta) + \frac{a_1}{n^{1/2}} - \frac{a_2}{n} \tag{A.3}$$

$$\begin{aligned}
&\frac{1}{n}[s_c^2 + (y_+ - X_R b_c)'(I - \frac{1}{n} X_R Q X_R')(y_* - X_R b_c)] \\
&= \sigma^2 f(1+\theta) + \frac{a_1}{n^{1/2}} - \frac{a_2 - \sigma^2}{n} + O_p(n^{-\frac{3}{2}})
\end{aligned} \tag{A.4}$$

where

$$\begin{aligned}
a_1 &= f^{1/2}[w_* + 2\beta'(u_* - u_R) - 2f^{1/2}\delta' Q(u_c + f^{1/2})u_R] \\
a_2 &= u_c'(Q - S_{cc}^{-1})u_c + 2f^{1/2}u_c' Q u_R + f u_R' Q u_R\,.
\end{aligned}$$

Thus the quantity h specified by (4.8) can be expressed as

$$\begin{aligned}
h &= 1 - \frac{1}{n f \sigma^2(1+\theta)}[\sigma^2 - \frac{w_c}{n^{1/2}} + O_p(n^{-1})] \\
&\quad [1 + \frac{a_1}{n^{1/2} f \sigma^2 (1+\theta)} + O_p(n^{-1})] \\
&= 1 - \frac{1}{n f(1+\theta)} + O_p(n^{-\frac{3}{2}})\,.
\end{aligned} \tag{A.5}$$

Using (A.1), (A.2) and (A.5) in (4.11, we get

$$\begin{aligned}
(\hat{\beta}_t - \beta) &= (1-t) f Q \delta + \frac{1}{n^{1/2}}[t Q_{cc}^{-1} u_c + (1-t) Q(u_c + f^{1/2} u_r))] \\
&\quad + \frac{t}{n(1+\theta)} Q\delta + O_p(n^{-\frac{2}{2}})\,.
\end{aligned} \tag{A.6}$$

It is easy to see that the bias vector of $\hat{\beta}_t$ to order $O(n^{-1})$ is

$$\begin{aligned}
B(\hat{\beta}_t) &= E(\hat{\beta}_t - \beta) \\
&= \sigma(1-t) f Q \phi + \frac{\sigma t}{n(1+\theta)} Q \phi \\
&= \sigma(1 - \alpha t) S X_R' Z \beta
\end{aligned} \tag{A.7}$$

where α and S are defined by (4.11) and (2.10) respectively.

Similarly, the mean squared error matrix of $\hat{\beta}_t$ to order $O(n^{-1})$ is given by

$$\begin{aligned} MSE(\hat{\beta}_t) &= E(\hat{\beta}_t - \beta)(\hat{\beta}_t - \beta)' \\ &= \sigma^2(1-t)^2 f^2 Q\phi\phi' Q + \frac{2\sigma^2(1-t)tf}{n(1+\theta)} Q\phi\phi' Q \\ &\quad + \frac{\sigma^2}{n}[t^2 Q_{cc}^{-1} + (1-t^2)Q] \\ &= \sigma^2(1-t)[1+(1-2\alpha)t] S X_R' Z\beta\beta' Z' X_R S \\ &\quad + \sigma^2[t^2 S_c + (1-t^2)S]\,. \end{aligned} \tag{A.8}$$

The results (A.7) and (A.8) are stated in the Theorem.

References

Little, R. J. A. (1992). Regression with missing X's: A review, *Journal of the American Statistical Association* **87**: 1227–1237.

Little, R. J. A. and Rubin, D. B. (1987). *Statistical Analysis with Missing Data*, Wiley, New York.

Rao, C. R. and Toutenburg, H. (1999). *Linear Models: Least Squares and Alternatives*, 2 edn, Springer, New York.

Toutenburg, H., Heumann, C., Fieger, A. and Park, S. H. (1995). Missing values in regression: Mixed and weighted mixed estimation, *in* V. Mammitzsch and H. Schneeweiß (eds), *Gauss Symposium*, de Gruyter, Berlin, pp. 289–301.